SFB 649 Discussion Paper 2005-036

Getting Used to Risks: Reference Dependence and Risk Inclusion

Astrid Matthey*

* Department of Economics, Humboldt-Universität zu Berlin and Department of Economics, Technische Universität Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Getting Used to Risks: Reference Dependence and Risk Inclusion

Astrid Matthey*

- Preliminary version -

May 13, 2005

Abstract

Experimental and field evidence show that people perceive and evaluate new risks differently from risks that are common. In particular, people get used to the presence of certain risks and become less eager to avoid them. We explain this observation by including risks in the reference states of individuals, which requires a more general concept of the reference state than has previously been considered in the literature.

We find two effects. First, the inclusion of the risk in the reference state changes its evaluation. A risk being present on the market induces a self-enforcing process of increasing acceptance of this risk without any new information becoming available or people's tastes changing. We term this the *risk inclusion effect*. Second, a risk with lower inherent (reference-independent) utility may be preferred to a risk with higher inherent utility if individuals are more used to accepting the former than the latter. This leads to inefficient decisions if habituation is seen as contributing less to welfare than inherent utility. Both effects have implications for the optimal regulation of risks.

JEL classification: D81, I18

Keywords: habituation, risk perception, reference-dependent preferences, WTA/WTP disparity, health risk

*Humboldt-Universität zu Berlin, Department of Economics, and Technische Universität Berlin, Department of Economics. Corresponding address: a_matthey@yahoo.com
I would like to thank Dorothea Kübler, Franz Hubert, Paul Heidhues, Daniel Krähmer, Julia Schmid and participants of the WZB and Humboldt University brown bag seminars for helpful comments and discussions. Financial support by the Deutsche Forschungsgemeinschaft through the SFB 649 is gratefully acknowledged.

1 Introduction

Experimental and field evidence show that people perceive and evaluate new risks differently from risks that are common. In particular, people get used to the presence of certain risks and become less eager to avoid them. We explain this observation by including risks in the reference states of individuals, which requires a more general concept of the reference state than has previously been considered in the literature.

We find two effects. First, the inclusion of the risk in the reference state changes its evaluation. A risk being present on the market induces a self-enforcing process of increasing acceptance of this risk without any new information becoming available or people's tastes changing. We term this the *risk inclusion effect.* Second, a risk with lower inherent (reference-independent) utility may be preferred to a risk with higher inherent utility if individuals are more used to accepting the former than the latter. This leads to inefficient decisions if habituation is seen as contributing less to welfare than inherent utility. It also offers a new explanation for the reduction of the disparity in willingness-to-pay (WTP) and willingness-to-accept (WTA) measures of value in repeated market-like settings.

Both effects have implications for the optimal regulation of risks. In particular, the risk inclusion effect induces a need to regulate the entry of new risks into the market with great care, since initially low acceptance levels may increase quickly, leading to increased risk production. Anecdotal evidence from the regulation of genetically modified Bt-corn in the US supports this view (see Bratspies, 2002). It shows that with lax regulatory requirements at the introduction of a risk, a subsequent tightening of the rules may be hard to implement.

That people's preferences are reference-dependent has been confirmed in many experiments. For instance, Kahneman and Tversky (1979) found that individuals evaluate prospects as losses and gains relative to a reference point rather than as absolute outcomes. As an implication of reference dependence, numerous experiments found an *endowment effect* (Thaler, 1980), according to which the value of an object increases once one possesses this object.

With reference-dependent preferences, the evaluation of prospects depends crucially on the definition of the reference state. Most models assume the reference state to be the status quo, that is, the person's current asset position or endowment (see Kahneman and Tversky, 1979; Kahneman, Knetsch, Thaler, 1990, 1991; Samuelson and Zeckhauser, 1988). An alternative approach emphasizes the role of expectations as reference point. Shalev (2000) takes the expected return on a portfolio as reference state for the evaluation of financial prospects. Köszegi and Rabin (2004) propose a person's recent belief or expectation about a future outcome as the reference point for this outcome. A third concept is presented by Sugden (2003), who includes acts that yield uncertain outcomes as possible reference states.

All these views have their merits, and it is hard to generally favor one over the other. However, one can easily reconcile them into a broader concept. Indeed, a conflict between the different concepts arises only if one interprets the status quo *narrowly*, namely as the physical endowments currently in one's possession (as is done in the status quo models). Instead, we suggest the interpretation of the status quo as the individual's current life situation. Then, it consists of much more than (state-dependent) physical endowments or expectations over future endowments. It includes all kinds of beliefs, perceptions, tastes, (sensory) feelings etc. Which of these components becomes relevant depends on the task, decision or situation that the individual evaluates. Regarding financial prospects, a person's current rate of return expectation may describe her status quo sufficiently well (as in Shalev, 2000), but other components of risk perception like familiarity, voluntariness etc. may also be relevant. If she is to evaluate the temperature in a room, her current feeling may constitute her status quo, but habituation to low or high temperatures may matter just as much. When evaluating health risks, the individual's current health state may have to be augmented with her risk perception to constitute her reference state, etc. Under such a broad concept, any feature of a person's current situation is part of her status quo and hence determines her reference state.

This concept is a generalization of the narrow status quo concepts used in the literature and incorporates those as special cases. However, it could be blamed to be too general to be applied. Indeed, for each evaluation it requires a decision about which of the components of the status quo is relevant. So are we back at square one? Not really. Our general concept makes clear that the various approaches that are discussed in the literature are merely applications of the same idea to different situations, and that the one "right" narrow definition of the reference state will likely not exist. In addition, it serves the important purpose of drawing attention to decisions where neither the physical-endowments approach nor the expected-outcome approach yield satisfying results, such that alternative concepts need to be developed.[1]

In this paper, we develop such a new concept for the evaluation of assets that cause non-monetary long-term risks, for instance, health risks. In particular, we explicitly include these risks in the reference state. There are three reasons why the concepts of the reference state that are proposed in the literature cannot be applied here. The first is that for the evaluation of such risks the *source* of the risk matters, and not only its outcomes. Second, the risk *itself* can become part of the reference state, rather than only its expected outcomes. And third, the inclusion of a risk in the reference state is a *gradual* process and cannot be described by a zero-one state variable as for physical endowments or expected purchase decisions.

Consider the first reason, namely that the source of a risk matters for its evaluation, rather than simply its outcome. Think, for example, of the risk from dying in a car accident vs. the risk from dying from a bomb attack. Both can lead to a sudden, violent death and can only partly be controlled (e.g., by driving carefully or rarely, and by avoiding crowded places). Hence, in terms of the outcomes they cause, the two risks can be interpreted as being roughly comparable. However, few people will consider them as similar. This is due to the fact that when people evaluate risks, they do not simply aggregate probabilities of (fatal) outcomes. Instead, the nature of the source of the risk matters for its evaluation (see Slovic (2001) for a comprehensive treatment of the factors that determine the percep-

[1] It should be clear from the discussion so far that these new concepts would not aim at *replacing* the existing approaches but at specifying the general approach for a class of prospects where these approaches cannot be applied.

tion of risks). Accordingly, when people assess risks, they tend to frame them narrowly, i.e., they evaluate each risk (source) in isolation rather than aggregating the outcomes of all risks in their portfolio.[2]

The second reason why a new concept is required is that risks *themselves* can become part of the reference state. This is motivated by the observation that among all the risks that individuals face, some are seemingly carried happily, while others, that appear to be comparable in their objective severity, are accepted with great reluctance. Examples for the first kind are the risks arising from commuting to work every day, eating one's usual diet or taking aspirin. For the second kind one could think of the risk from terrorist attacks or the risk from eating genetically modified (gm) food for European consumers. We claim that the difference in the evaluation of these risks arises (at least partly) from the first kind being included in individuals' reference states (known risks), and the second kind being excluded (new risks). Health risks are special in that for many such risks, outcomes are expected to become apparent only several years after the risk was accepted. During this time, i.e., as long as one does not have an accident, does not develop a food-related illness etc., outcomes cannot become included in the reference state. Hence, if habituation takes place, individuals get used to the mere presence of a risk, rather than to its outcomes.

Experimental evidence supports this claim. In experiments by Thaler (1980) and Viscusi, Magat and Huber (1987) an endowment effect was found for the acceptance of risks. In Thaler's experiment, subjects were first asked how much they would be willing to pay in order to cure a disease they had been exposed to, which had a 0.001 risk of death. Then, they were asked how much they would have to be paid in order to be exposed to this disease willingly, i.e., accept an additional risk of death of 0.001. Thaler finds that most people were willing to pay much less in the first case then they demanded in the second. This means that once a risk is accepted, individuals' willingness to pay for eliminating it is reduced compared to the willingness to accept payment for taking the risk before one first did so. Hence, the endowment effect, which arises from a change in the reference point is present also for risks.

Viscusi, Magat, Huber (1987) obtained a similar result in an experiment where shoppers were asked about their willingness to pay for the reduction of the existing risk from standard insecticides vs. the price they would demand in order to accept an increase in this risk.[3] In the latter case, 77% of the subjects bluntly refused to buy the product with the increased risk at any positive price. Those that did accept the riskier product demanded a premium 6 to 10 times higher than that for risk reductions. An interesting detail here is that the subjects were not actually endowed with either of the insecticides. The only difference between the two products was that in the first case, subjects would have to pay for a reduction of the risk from *standard* insecticides, while in the second case they where asked to

[2]Taken literally, it is more to the point to say that the vectors of relevant risk characteristics are so complex, i.e., they include so many different dimensions to describe a risk, that two risks are never comparable in all dimensions. Accordingly, if utility is non-separable in risk dimensions, risks have to be evaluated in isolation.

[3]The two designs differ in that Thaler considers a *new* risk, while Viscusi, Magat, Huber consider an increase in a *known* risk. However, this increase seems to lead to a qualitative difference in the perception of the risk, such that the two cases are indeed comparable.

accept a risk level *above* the normal standard (which represents the qualitatively *new* risk). Hence, (expected) physical possession did not matter, while a risk being *standard* or *common* did matter.[4]

These experiments show that subjects have a tendency to pay less to avoid an existing risk than they demand to accept a new risk. The interpretation of the endowment effect indicates that this is due to some risks being included in the reference state and others being excluded. An important question then is how the process of risk inclusion works. This leads to the third reason for developing a new model. In the two experiments described above (and in most literature on the endowment effect and reference-dependent preferences in general), only the two extreme cases are considered: complete reference state inclusion and complete exclusion. This binary perception of the reference state stems from its presentation as either a narrow physical-endowment status quo (to have or not to have) or the expectation over a future endowment (e.g., to buy or not to buy).[5] If, instead, one interprets the reference state to include concepts like habituation, it cannot be expected to change ad hoc, i.e., to be describable by a binary $\{0, 1\}$ variable.[6] Rather, reference state changes must be expected to happen gradually over time.

The notion of a gradual reference state change is not new to the economic literature; it was already widely applied in models of habit formation (Duesenberry, 1949; Pollak, 1970; Ryder and Heal, 1973).[7] With habit formation, a person's consumption level today and/or the consumption level of others influences her future habit level, which constitutes the reference state and affects her preferences (see, e.g., Abel, 1990; Constantinides, 1990; Wathieu, 1999). There are, however, substantial differences between our model and models of habit formation. The latter refer to *realized monetary* outcomes, i.e., to a narrow status quo definition, which does not include risks. In addition, most of the models apply habit formation to financial decision making (Campbell and Cochrane, 1997; Gomes and Michaelides, 2003; Polkovnichenko, 2003), and consider aggregated consumption instead of disintegrating it into particular assets. With such an approach, the sources of the risk cannot be taken into consideration. Hence, individual risks cannot be included in the reference state. Yet, the ideas underlying habit formation and risk inclusion are related: persistence of a state of nature leads people to get used to this state, which gradually changes their evaluation of this state and other states related to it. Hence, when specifying the risk inclusion process we are able to draw on some results of this literature.

Assuming a gradual risk inclusion process, its endpoint is naturally hard to determine. Consider again the experiment by Viscusi, Magat and Huber (1987).

[4]The experiments also show that the endowment effect can be observed for non-monetary risks, a view that is supported by the work of Breyer and Fuchs (1982). The authors conducted a questionnaire survey in US hospitals and found the main implications of reference-dependent preferences confirmed in the context of health risks.

[5]Shalev's (2000) expected return is somewhat of an exception. However, his concept is markedly different from ours in that a change in the reference state in his model requires new information on the investment (its outcomes or outcome distribution) or a change in the individual's utility function.

[6]Köszegi and Rabin (2004) do not specify the process of expectation building, leaving room for gradual adjustment. However, the presentation of expectations in their model does not imply a gradual approach.

[7]See also the literature on social comparison, e.g., Festinger (1954) and Wood (1989).

How can one know whether the standard insecticide is 80% or 100% included in the reference state of individuals? In practice, one cannot. This means that one cannot know whether the authors found a *full* or just a *partial* endowment effect.

What one can infer relatively safely is that they found *an* endowment effect, and hence there has to be a process at work that makes a formerly new risk an endowed risk. This is what matters for our argument. In the model, we will assume the risk inclusion process to begin when the first individual accepts the risk and to end when all individuals accept it, a state that will hardly obtain in practice. Nevertheless, an endowment effect will be observable whenever there is a difference in people's habituation to two otherwise similar risks. Regarding the experiment by Viscusi, Magat and Huber, we would expect the difference in WTP/WTA to decrease if the new insecticide with the increased risk was available on the market for some time.

The remainder of the paper is organized as follows. In section 2 we develop our concept of the reference state, from which we derive the risk inclusion effect in section 3. The individual's utility function is specified and an explanation of the WTA/WTP difference given in section 4. Section 5 applies the results to the individual's risk acceptance decision. In section 6 we analyze the implications for risk regulation. Field evidence is presented in section 7. Section 8 concludes and gives a brief outlook on future research.

2 Reference state determination

We focus on non-monetary long-term risks that do not cause adverse effects immediately but only some time after they were accepted. An obvious application are health risks caused by food, life style, driving behavior etc. For concreteness, we will refer to these risks throughout the paper, but the application of the model is not limited to them. We shall call the source of the risk an 'asset', which is meant to include goods, actions etc.

The risky asset x induces a set of possible outcome vectors, $x = \{x_1, ..., x_i, ..., x_I\}$. Each x_i should be thought of as a vector of health dimensions. The set of possible outcome vectors is discrete and bounded with $x_i \in [x_1, 0]$. This means that in the best case the risk does not cause an effect in any dimension (unchanged health state), while in the worst case it causes outcome x_1. Without loss of generality, let us assume that outcomes can be ordered according to their severity: $x_1 \prec .. \prec x_i \prec x_{i+1} \prec .. \preceq 0$. Individuals have a common prior distribution over the set of possible outcomes, denoted $p(x)$.

There is a population of individuals, which is characterized by its members being able to observe each others' risk decisions (e.g., through the media, personal relations etc.). Let $t = 0$ be the time when the risky asset is first chosen by a member of the population. We assume that the risk enters people's reference states only if at least one member of the population accepts it. This means that up to $t = 0$, the risk is not included in any individual's reference state: $\hat{g}_t = 0$ for $t \leq 0$, where $\hat{g} \in [0, 1]$ denotes the degree of reference state inclusion of the risk.

The process of risk inclusion depends, first, on the individual's own past decisions, denoted by the binary variable $h_t \in \{0, 1\}$, where $h_t = 1$ if the risk was accepted and $h_t = 0$ otherwise. Second, it depends on the aggregated decisions of the other members of the population, where $H_t \in [0,1]$ denotes the share of people that accept the risk in t. The degree of reference state inclusion in period t is then given as

$$\hat{g}_t = \phi\hat{g}_{t-1} + (1-\phi)(\alpha h_{t-1} + (1-\alpha)H_{t-1}) \qquad \alpha, \phi \in [0,1]. \tag{1}$$

ϕ and α are individual parameters. ϕ determines the speed of reference state inclusion. α determines the relevance of an individual's own decision vs. the aggregated decisions of others for her degree of risk inclusion. Note that for $\phi = 0$ and $\alpha = 1$, we recover the common endowment effect where the reference state is equal to one's narrow status quo or endowment last period.

This specification of the degree of reference state inclusion is related to the habit level in models of habit formation. The inclusion of one's own decisions reflects *internal* habit formation as, for example, in Sundaresan (1989) and Constantinides (1990). This concept is recovered from (1) for $\phi > 0$ and $\alpha = 1$. The relevance of other people's decisions reflects social comparison preferences or *external* habit formation (see, e.g., Abel, 1990; Campbell and Cochrane, 1997), which is obtained from (1) for $\phi > 0$ and $\alpha = 0$. Hence, through the calibration of ϕ and α, (1) can account for the most widely used concepts of the reference state. We will show that our results hold as long as $\phi < 1$ and $\alpha < 1$, i.e., as long as the aggregated decisions of other members of the population have an impact on the reference state of the individual.

As a minimum rationality condition, we assume that subjects make consistent decisions in the sense that they follow the same decision concept each period. This means that changes in decisions arise only from changes in the parameters of the decision program. Hence, once an individual decided to accept a risk, she will continue to accept this risk until some parameter change induces her to avoid it. As we will show below, risk acceptance is monotone increasing in the degree of reference state inclusion, meaning that once an individual accepts a risk, without changes in the exogenous parameters she will accept the risk in all future periods: $h_t = 1 \Rightarrow h_s = 1 \; \forall s \geq t$. When determining the process of risk acceptance, i.e., the share of individuals accepting the risk, this monotonicity allows us to consider at time t only those individuals for whom $h_{t-1} = 0$. Hence, the degree of reference state inclusion can be reduced to

$$g_t = \phi g_{t-1} + (1-\phi)(1-\alpha)H_{t-1} \qquad \alpha, \phi \in [0,1]. \tag{2}$$

The reference state r of the risk is a vector that consists of the same health dimensions as the outcomes x_i. In contrast, g_t is simply a number between zero and one. Consider first the two extreme cases. Complete risk exclusion means that the degree of reference state inclusion is zero: $g_t = 0$. Accordingly, the reference state regarding the risk under consideration is "no health effect": $r = \mathbf{0}$. With complete risk inclusion, $g_t = 1$. Then, the risk itself serves as reference: $r = x$. Partial risk inclusion, $g_t \in (0,1)$, means that the risk is included in the reference state to less than full extent. We do not claim to know what "inclusion to less than full extent" means psychologically. The only assumption necessary for our argument is that for $x \neq 0$ and $g < g'$: $gx|r \succ g'x|r \; \forall r$ with $g_t x = \{g_t x_1, .., g_t x_I\}$.

This means that less of the risk (less severe outcomes) is preferred to more of the risk for all reference states.[8] However, in order to make our argument clear and keep the model as simple as possible, we specify a process of reference state change that is linear in the degree of risk inclusion g_t in each period t:

$$r_t = (1 - g_t)\mathbf{0} + g_t x = g_t x \tag{3}$$

where $x = \{x_1, .., x_I\}$ as before. Remember that r_t does not denote the complete reference state of the individual but only the part that focuses on the particular risk under consideration.[9]

3 Risk inclusion effect

We assume that the preferences of all individuals under consideration satisfy the axioms necessary for a subjective expected utility (SEU) representation of reference-dependent preferences to exist. Sugden (2003) provides such an axiomatization where he extends the original Savage (1954) axioms to account for reference dependence. In particular, the SEU representation exists if there is a unique probability measure p, and a relative value function v, such that for all Savage-acts j, k, l: $j \succeq k|l \iff E_p[v(j|l) - v(k|l)] \geq 0$ (Sugden, 2003, p. 175), i.e., if an act j is weakly preferred to an act k when both are viewed from l, if and only if the expected difference in relative values of j under l and k under l is non-negative.

The relative value function is defined as a finitely valued function $v : X \times X \to \mathbf{R}$, where X is the set of consequences or outcomes, and

A1: $v(x|x) = 0$ for all x.

Following Kahneman and Tversky (1979) we add two additional assumptions:

A2: $v(x|r)$ is continuous for all x and twice differentiable for $x \neq r$;

A3: $v'(x|r) > 0$ for all x given r.

$v(x|r)$ is to be interpreted as a measure of the value of x, relative to r, when both are viewed from the reference point r.

$A1$ shows that v represents only *relative* values, i.e., values of changes from a certain reference state. $A2$ is a technical assumption (which allows for loss aversion, i.e., a "kink" at $x = r$). $A3$ means that v is increasing in x over the range of relevant values. This is not a straightforward assumption since, in contrast to most other models of reference-dependent preferences, we do not consider monetary outcomes here. We do, however, assume outcomes to be ordered with increasing preferability. $v(x|r)$ can thus be interpreted as to be increasing in the index of x_i.

From the above specifications, the effect of risk inclusion on the relative value of the risk follows immediately:

[8] Throughout the paper, $x|y$ will be taken to mean "x viewed from y".

[9] Over the life time of the risk, two components of the reference state are relevant. The first is the one that we consider in this paper, namely the reference state regarding the risk that the individual accepts. The second component is the reference state regarding the individual's actual health state. It becomes relevant only if the adverse health effects actually occur. Then, the reference state must be expected to be the health state immediately before the occurrence of these effects, such that they will always be perceived as losses. This second component is not considered in this paper.

Proposition 1 *If $v(x|r)$ is a relative value function that satisfies $A1 - A3$, then the following holds:*
For all $x \neq 0$ and g, g' such that $g < g'$,

$$v(x|gx) < v(x|g'x). \tag{4}$$

The proof is obvious. For $g < g'$ and $x \neq 0 \Rightarrow gx \succ g'x$: less risk is preferred to more. Proposition 1 then follows directly for increasing relative value functions: a given risk or vector of possible outcomes under a lower reference point is preferred to the same risk under a higher reference point. Similarly, it can be shown that for all $x_i \neq 0$ and g, g' such that $g < g'$, $v(x_i|gx_i) < v(x_i|g'x_i)$. The proposition holds independently of how the health dimensions of x_i are aggregated into $v(x|gx)$, as long as an increase in value in one dimension does not lead to a decrease in overall value. In particular, it does not require additive separability of the value function across dimensions. To summarize, Proposition 1 shows that the value that an individual assigns to a given risk increases with an increasing degree of risk inclusion, without any new information becoming available.

Assuming that probabilities are linear[10] we obtain the value of the risky asset as

$$V(x, g_t) = \sum_i v(x_i|g_t x_i)p(x_i). \tag{5}$$

Since $v(x_i|gx_i)$ is increasing in g, and as $p(x)$ is independent of g, $V(x,g)$ is increasing in g. Hence, we have established that for any increasing relative value function the value that an individual assigns to a risky asset is increasing in the degree of reference state inclusion of the risk. We shall term this result the *risk inclusion effect*.

4 Individual utility

4.1 Utility function

This section considers the implications of the risk inclusion effect for the total utility of the individual. The utility from accepting the risky asset has two components. First, it yields present utility $u_p(x)$ immediately after consumption (e.g., not feeling hungry anymore after eating). Second, it gives rise to future utility $u_f(x)$ by affecting the individual's health some time after consumption. Future utility $u_f(x)$ again has two parts: the inherent utility from the health risk, $u(x)$, which is reference-independent, and the relative value $v(x|r)$ which depends on the reference state.[11] We obtain the overall utility from consuming asset x as

$$\begin{aligned} U(x,r) &= u_p(x) + u_f(x) \\ &= u_p(x) + u(x) + v(x|r) \end{aligned} \tag{6}$$

[10]For theory and evidence on non-linear probability weights, see e.g., Kahneman and Tversky (1979), Wu and Gonzalez (1999) and Starmer (2000). The effect of risk inclusion with probability weights will be considered in a companion paper.

[11]$u(x)$ and $v(x|r)$ in our model are related to the "attitudes to end states" and "attitudes to gain and loss" in Sugden (2003) or the "consumption" and "loss-gain" utility in Köszegi and Rabin (2004). The difference is, however, that we do not consider realized outcomes but accepted risks.

where r is defined according to (3) and we abstract from discounting.

Consider two risky assets x and z that induce reference-independent utilities $\bar{U}(x) = u_p(x) + u(x)$ and $\bar{U}(z) = u_p(z) + u(z)$ and relative values $v(x|g_x x)$ and $v(z|g_z z)$. Specify an increase ϵ in the utility of one asset and two degrees of reference state inclusion g_x and g_z. Then the following proposition holds:

Proposition 2 *For two risky assets x and z such that $\bar{U}(x) = \bar{U}(z)$, and $v(x|g_x x) = v(z|g_z z)$ for $g_x = g_z$:*

If $g_x > g_z \Rightarrow \exists \epsilon > 0$ such that $\bar{U}(x) + \epsilon = \bar{U}(z)$ and $U(x, r) > U(z, r)$, i.e., $x \succ z$.

The proof follows directly from (6) and Proposition 1. Proposition 2 shows that for two risky assets between which an individual with either reference-independent or reference-dependent preferences is originally indifferent, one can always add a positive amount ϵ to the reference-independent utility of one asset such that this amount can be offset by a higher degree of reference state inclusion. An individual with reference-dependent preferences will then prefer the asset with the lower reference-independent utility to the asset with the higher reference-independent utility.

The result of Proposition 2 implies inefficient decisions of individuals if $v(x|r)$ is not perceived as contributing towards welfare to the same extent as does reference-independent utility. Different levels of habituation to risks can then lead to welfare-decreasing consumer decisions. See section 6 for a more comprehensive interpretation of this result.

From $v(x|x) = 0$ and $v(x|r) \leq 0$ for $x \leq r$ (see $A1$ and $A3$), and (4) we obtain that $lim_{g \to 1} v(x|gx) = 0$. Hence,

$$lim_{g \to 1} U(x, r) = u_p(x) + u(x). \tag{7}$$

This expression shows that with increasing inclusion of the risk in the reference state the utility from the risky asset converges towards its inherent utility. In other words, when the risk gets fully included in the reference state, an individual with reference-dependent preferences and an individual with a classical von Neumann-Morgenstern utility function derive the same utility from accepting the risk (provided that they have identical inherent utility functions $u_p(\cdot)$ and $u(\cdot)$). However, this does not mean that the reference-dependent preferences of the individual become increasingly reference-independent. Rather, what was perceived as a loss from the original reference point (and caused reference dependence to become effective) is not considered a loss anymore.

The effect from (7) offers a new explanation for the observation that the difference between WTP and WTA diminishes when preferences are elicited repeatedly in a market-like setting, which we discuss in the next section.

4.2 WTA/WTP disparity and risk inclusion

In this section, we will argue that the observed decrease in the WTA/WTP disparity can by explained with the risk inclusion effect. Since Knetsch and Sinden

(1984) published the results of a series of experiments showing significant differences between willingness-to-pay and willingness-to-accept measures of value, other experiments have been conducted to show that this anomaly is greatly reduced in repeated market-like settings (Coursey, Hovis and Schulze, 1987; Shogren et al., 2001; Loomes, Starmer and Sugden, 2003). Without giving a full overview over the literature, we will reconsider some of the experiments and their interpretation in light of our model.

Coursey et al. (1987) ran repeated Vickrey-like auctions of the obligation to taste a bitter substance (sucrose octa-acetate or SOA), while Loomes et al. (2003) ran similar auctions of vouchers and lotteries. The results of both experiments show a significant disparity between WTA and WTP measures of value in the initial rounds, but a reduction of this disparity in later rounds. From their results, Coursey et al. infer that any difference in elicited values is merely due to mistakes made by the subjects, rather than to differences in the "true" preferences. Through repeated auctions, individuals learn both their true preferences and that revealing their true value is their dominant strategy. Similarly, Loomes et al. conclude that "market experience *does* tend to erode whatever causal factors generate the tendency for WTA to be systematically greater than WTP" and that "the factor being eroded is something other than, or additional to, strategic bias." (p. C165). The authors offer the explanation that loss averse individuals "by participating in markets in which endowments are routinely sold, [..] become more familiar with, and so less averse to, the idea of selling" (p. C166).[12]

This latter point, which the authors do not elaborate on, hints towards the argument of our model. Consider the situation subjects are faced with in the two experiments. The results of the first auction rounds may be different from the subject's initial status quo. If this is the case, the reference state will adjust (fully or partially) towards this new state in later rounds. Under a new reference state, the relative value of prospects changes, leading to different WTA/WTP measures in the final rounds. The convergence of WTP and WTA would then be observed also for subjects whose preferences are reference-dependent. The "correction of mistakes" or "erosion of causal factors" would not be necessary to explain the observed behavior.[13]

Note that the described experiments correspond with our model in another important feature, namely that subjects do not obtain feedback on the outcomes

[12] An alternative experiment is reported by List (2003). He analyzes the trading behavior of traders vs. visitors at a US sports card fare. He finds fairly large WTA/WTP differences for relatively inexperienced consumers, but these differences disappear for experienced consumers and traders. However, the line of causality is not clear in his experiment. Rather than concluding like List that by trading more frequently, people's behavior converges towards reference independence, one could also claim that people whose preferences are relatively reference-independent (or who are inclined to trading or gambling), tend to trade more frequently and will have come to the fare in a "gambling mood" (as either traders or consumers).

[13] The same argument applies to the experiment of Shogren et al. (2001), who tested for the effect of substitutability on the WTA/WTP measures in repeated auction settings.
Our model can also explain the difference in WTA and WTP trends that Coursey et al. found, namely that the WTA measures decreased, but the WTP measures did not increase. For WTA, reference state adoption means changing the notion of "tasting SOA" from a "loss" into "not-so-clearly a loss". For WTP, on the other hand, "not tasting SOA" is changed from a "gain" into "not-so-clearly a gain". If subjects are loss averse, preferences are more strongly affected in the first case than in the second.

of their decisions. In both experiments, the subjects did not know when bidding whether the current round would be decisive for their final payoffs. For instance, the losers of an auction did not necessarily have to taste SOA. Hence, if a subject in one round found herself in a state different from her (prior) status quo, the sole *possibility* that this round may turn out to be decisive would have led to a change in her reference state. The observed convergence of WTP and WTA measures would then be fully in line with the predictions of our model.[14]

5 Risk acceptance decision

5.1 Individual risk acceptance decision

We analyze the individual's risk acceptance decision in a simple two-period model (see figure 1). Consider first the monetary flows. In period 1, she earns income Y_1, which she can invest into consumption c_1 or in a safe investment B. In period 2, she earns income Y_2, receives $B(1+i)$ and consumes all wealth, c_2. If the individual performs mental accounting (see, e.g., Thaler, 1999), i.e., she aggregates expenditures with the same intended purpose, Y and B are to be interpreted as those parts of income and investment that are assigned to the same mental account as x and the substitute safe good (e.g., food). For simplicity, we assume the interest rate i to be the same for saving ($B > 0$) and borrowing ($B < 0$). i is the real interest rate between periods 1 and 2, i.e., it is unknown in period 1. Y_2 and i impose a monetary risk on the individual, that is, on her consumption level in period 2.

The individual chooses between the risky asset x and a safe but more expensive comparable asset y. Present utility is equal for both assets and normalized to zero: $u_p(x) = u_p(y) = 0$. Since y is a safe asset which does not cause adverse health effects its future utility is $u_f(y) = 0$. This yields utilities for the two assets of

$$\begin{aligned} U(x,r) &= u(x) + v(x|r) && \text{for the risky asset} \\ U(y,r) &= 0 && \text{for the safe asset.} \end{aligned}$$

The risky asset sells at price P_x and the safe asset at price P_y, where $P_y = P_x + \pi$. For ease of notation, we set $P_x = 0$, and hence $P_y = \pi$. The decision whether to purchase the risky asset (accept the risk and pay the lower price) or the safe asset (avoid the risk and pay the higher price π) is made in period 1. Hence, period 1 is identical with the time t when the degree of risk inclusion is evaluated.[15] The health effects of the risk occur in period 2. Note that the time between periods 1 and 2 is far longer than between t and $t+1$. The parameter ξ indicates whether an individual accepts the risk ($\xi = 0$) or not ($\xi = 1$).

[14] Formally, the definitions of h and H have to be changed slightly. Since "true" risk acceptance is unknown when bidding, h is not a simple binary variable anymore. Instead, expected values are taken for both h and H such that $E(h) \in [0,1]$ and $E(H) \in [0,1]$. Risk inclusion is then determined by $\hat{g}^*_t = \phi\hat{g}^*_{t-1} + (1-\phi)(\alpha E(h_{t-1}) + (1-\alpha)E(H_{t-1}))$, see (1). Despite the different formal representation, the argument is identical with the one for g_t, such that we will not consider $\hat{g}^*_t$ in detail.

[15] Accordingly, each t at which the individual evaluates the risk induces its "own" periods 1 and 2.

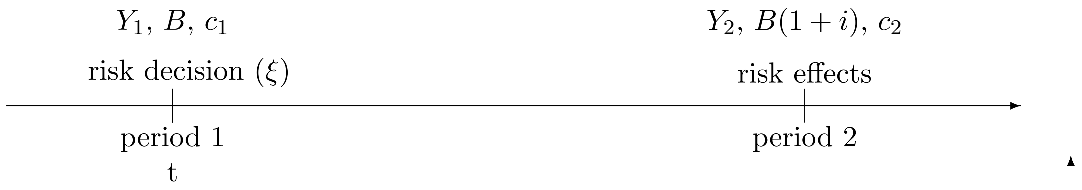

Figure 1: Timing

Similarly to the health risk, the evaluation of the monetary risk depends on the individual's reference consumption r_c. Hence, the utility from consumption of monetary goods also has a reference-independent and a reference-dependent component. However, for monetary risks both components arise immediately after consumption: $U_c(c, r_c) = u_c(c) + v_c(c|r_c)$.[16] We assume that the reference point for future consumption is formed from expectations over future consumption possibilities (see Shalev, 2000, and Köszegi and Rabin, 2004). This means that the reference point in period 2, r_{c_2}, is determined by the individual's expectations over c_2, i.e., over Y_2 and i. Hence, in period 1 $E(c_2)$ is neither viewed as a gain nor as a loss: $E(v_c(c_2|r_{c_2})) = 0$.[17] This implicitly assumes that individuals do not anticipate any endowment or habit formation effects on consumption between periods 1 and 2. We make this assumption for the following reason. The time between the risk acceptance decision and the expected health effects from the risk may be substantial, such that several consumption periods may lay in between the "risk relevant" periods 1 and 2. Hence, calculating habit effects would require detailed knowledge of the future consumption path. $E(U_c(c_2, r_{c_2}))$ would only be affected by habit formation if individuals had an expectation over this path and over the effects that it will have on their reference points in period 2. We will assume that individuals are not able or willing to form such sophisticated expectations (see also Munro and Sugden, 2003).

The individual's optimal risk acceptance decision is then obtained from the program

$$max_{B,\xi} \quad U_c(c_1, r_{c_1}) \quad + \quad \beta\{E[U_c(c_2, r_{c_2}) + (1-\xi)U(x, r)]\} \tag{8}$$

$$\begin{aligned} s.t. \qquad c_1 &= Y_1 - B - \xi\pi \geq 0 \\ c_2 &= Y_2 + B(1+i) \\ U(x, r) &= u(x) + v(x|gx) \\ U_c(c, r_c) &= u_c(c) + v_c(c|r_c) \end{aligned}$$

β is the discount factor between periods 1 and 2. In analogy to representations

[16] That there is no lagged utility from overall consumption is a simplifying assumption, which implies that neither c_1 nor c_2 are expected to induce any long-term health risks. Portfolio effects from accepting several risks with possible long-term effects are considered in a companion paper.

[17] The expectation operator throughout the paper denotes expectations in period 1. The index is therefore suppressed.

of reference-independent preferences, in (8) we assume utility to be separable in monetary and health goods.

For simplicity, we solve the optimization program separately for $\xi = 0$ and $\xi = 1$, i.e., we do not consider intermediate values $0 < \xi < 1$. This means that we assume the risk decision to be binary: the risk is either fully accepted ("eat gm-tomatoes") or completely avoided ("do not eat gm-tomatoes"), but not partly accepted and avoided ("eat gm-tomatoes during the week but not on weekends"). This seems a reasonable assumption when interpreting the risky asset as a consumption good or policy measure (support or do not support). From this we obtain the optimal saving decisions and the maximum utility that the individual can receive in both cases.

With $E_t[U(x,r)] = \sum_i [u(x_i) + v(x_i|r)]p(x_i) = \sum_i u(x_i)p(x_i) + V(x, g_t)$, the overall utility from (8) can be written as

$$\begin{aligned}\mathcal{U}(B,\xi) = U_c(Y_1 - B - \xi\pi, r_{c_1}) &+ \beta E_t[U_c(Y_2 + B(1+i), r_{c_2})] \\ &+ \beta(1-\xi)(\sum_i u(x_i)p(x_i) + V(x, g_t))\end{aligned}$$

The individual's utility for $\xi = 0$ and $\xi = 1$ is given by

$$\begin{aligned}\mathcal{U}(B,\xi = 0) = U_c(Y_1 - B, r_{c_1}) + \beta E_t[U_c(Y_2 + B(1+i), r_{c_2,\xi=0})] \\ + \beta(\sum_i u(x_i)p(x_i) + V(x, g_t)) \qquad (9)\end{aligned}$$

$$\mathcal{U}(B,\xi = 1) = U_c(Y_1 - B - \pi, r_{c_1}) + \beta E_t[U_c(Y_2 + B(1+i), r_{c_2,\xi=1})] \qquad (10)$$

The optimal investments into the bond are determined by the first order conditions

$$U_c'(Y_1 - B_{\xi=0}, r_{c_1}) = \beta E[(1+i)u_c'(Y_2 + B_{\xi=0}(1+i))] \qquad (11)$$
$$U_c'(Y_1 - B_{\xi=1} - \pi, r_{c_1}) = \beta E[(1+i)u_c'(Y_2 + B_{\xi=1}(1+i))] \quad , \qquad (12)$$

where we use the fact that $E(v_c(c_2|r_{c_2})) = 0$, i.e., $E[U_c'(Y_2 + B_\xi(1+i), r_{c_2})] = E(u_c'(Y_2 + B_\xi(1+i)))$. Due to the separability of the utility function, the optimal amount $B^*_{\xi=0}$ when the risk is accepted is independent of the expected effect from the risk. The optimal amount $B^*_{\xi=1}$ when the risk is avoided depends on the price π of risk avoidance.

To derive the optimal investments B^*, the first order conditions (11) and (12) need to be considered for two different cases: i) c_1 is evaluated as a gain, i.e., it is higher than expected; ii) c_1 is evaluated as a loss, i.e., it is lower than expected.

For risk averse or risk neutral individuals, standard utility $u_c(c)$ is increasing in c and weakly concave over its entire range.[18] $v_c(c|r_c)$ has also been found to be concave in the range of gains. In the range of losses, Kahneman and Tversky (1979) found $v_c(c|r_c)$ to be convex, while the habit formation literature assumes it to be concave. Hence, we have that $U_c(c, r_c)$ is concave for gains, but may be concave or convex for losses. The curvature of U_c for losses if v_c is convex depends on the relative curvature of u_c and v_c at consumption levels c and r_c.

$E(U_c(c_2, r_{c_2})) = E(u_c(c_2))$, i.e., it is concave for all consumption levels c_2. $U_c(c_1, r_{c_1})$ is determined jointly by u_c and v_c. u_c is concave everywhere and steeper for smaller

[18]We do not consider risk loving individuals.

values ("losses" in the sense of v_c.). If v_c is also concave for losses, $U_c(c_1, r_{c_1})$ is concave for c_1 over its entire range. Let us first consider this case. Fix a $B^*_{\xi=0}$ such that (11) is satisfied. For $B_{\xi=1} = B^*_{\xi=0}$ the left hand side of (12) exceeds the right hand side. In order to satisfy (12), $B_{\xi=1}$ has to decrease, such that the left hand term decreases while the right hand term increases. As the result, we have that

$$B^*_{\xi=1} < B^*_{\xi=0} \tag{13}$$

$$\text{but} \quad B^*_{\xi=1} + \pi > B^*_{\xi=0} \quad . \tag{14}$$

If an individual accepts the risk, she saves more or borrows less than if she avoids it, but the difference is less than the price premium π.

Consider now the case when v_c is convex for losses, i.e., we are in the context of loss aversion with diminishing sensitivity. For loss averse individuals, v_c has been found to be steeper for losses than for gains. Kahneman and Tversky (1979) have estimated that $v_c(x < r|r)$ is steeper than $v_c(x > r|r)$ by a factor 2 to 2.5. Since u_c is also steeper for smaller x, the aggregated function $U_c(c_1, r_{c_1})$ is steeper for losses by a factor exceeding 2.5. Hence, since $E[u_c(c_2)]$ is reference-independent, for most realizations of Y_2 and i optimization will lead to the choice of B^* such that c_1 is considered as a gain. This is because due to the steepness of the utility function for losses, for $U_c(c_1 < r_{c_1}, r_{c_1})$ and $r_{c_1} - c_1$ not too large, utility U_c is increasing much faster for increasing c_1 than it is decreasing for diminishing c_2. B^* would then be chosen small enough such that $c_1 > r_{c_1}$. In this case, $U_c(c_1, r_{c_1})$ is again concave in the relevant range and the above argumentation applies, leading to the results (13) and (14).

By the same argument, if c_1 is considered a loss before paying π, after the payment of π the adjustment of B^* will in most cases move c_1 "back" towards smaller losses, again leading to results (13) and (14). Only for extreme distributions of c_1 and c_2 and strongly convex relative value functions (rapidly diminishing sensitivity) will individuals find it optimal to accumulate huge losses in period 1. Such extreme constellations were required in order for the payment of π to induce a move towards larger losses, i.e., $B^*_{\xi=1} > B^*_{\xi=0}$. In what follows we will abstract from these extreme cases and assume (13) and (14) to hold.

Applying $B^*_{\xi=0}$ and $B^*_{\xi=1}$ to (9) and (10), we obtain the following condition for the avoidance of the risk:

$$\beta[\sum_i u(x_i)p(x_i) + V(x, g_t)] < \underbrace{U_c(Y_1 - B^*_{\xi=1} - \pi, r_{c_1}) - U_c(Y_1 - B^*_{\xi=0}, r_{c_1})}_{:=a} + \\ + \beta E[\underbrace{u_c(Y_2 + B^*_{\xi=1}(1+i)) - u_c(Y_2 + B^*_{\xi=0}(1+i))}_{:=b}] \tag{15}$$

Intuitively, the term a measures the individual's decrease in utility from lower consumption in period 1 due to the monetary loss π, which is partly offset by the difference in investments, $B^*_{\xi=0} - B^*_{\xi=1}$. Term b measures the decrease in utility from lower consumption in period 2 due to having saved less or borrowed more in period 1. From (14) and (13) we obtain that $a < 0$ and $b < 0$, respectively. Hence, if the discounted value of the risk exceeds in absolute terms the decrease in (discounted) expected value from monetary consumption in periods 1 and 2, it is optimal for the individual to avoid the risk. (15) implicitly determines a degree of risk inclusion g^* for each individual such that for $g_t = g^*$ she is indifferent between accepting and avoiding the risk.

If the expected value of the risk, $V(x|g_t x)$, increases due to an increase in g_t, the condition for the avoidance of the risk is tightened, i.e., people's willingness to pay for risk avoidance is reduced. This results in a (weakly) monotone increase in the share of risk acceptance, which justifies the monotonicity assumption we made in section 2. If more people accept the risk, the reference state increases for all individuals with $\phi \neq 1$ and $\alpha \neq 1$,[19] leading to a further increase in $V(x|g_t x)$. Hence, the process of increasing aggregated risk acceptance is *self-enforcing*, that is, it does not require that the population obtains any new information regarding the risk. Consequently, as implied by Proposition 2, acceptance of two objectively identical risks may differ widely depending on how long the risk inclusion process is already at work. We discuss the implications of this result in section 6.

Remark 1 *In the optimization program (8) it is assumed that the individual does not consider the price π of avoiding the risk in isolation but within its mental account. Whether this is actually the case or not depends, unsurprisingly, on the reference point for the expenditure. In our model, the risk is new to the individual, the higher price for the known, safe good constitutes the status quo. Hence, the formalization in (8) correctly captures this case. One could, however, imagine, that just as with the risk, some habituation to the lower price of the risky asset takes place, even if the individual does not purchase it. The reference price regarding the good would then move from π towards 0, the normalized price of x. This would result in a perceived (partial) loss when paying the higher price π, moving the preferences further towards acceptance of the risky good.*

5.2 Aggregated risk acceptance decision

If new information becomes available, we have to distinguish between information that increases $V(x,r)$ (favorable information) and such that decreases $V(x,r)$ (unfavorable information). Favorable information leads to a downward shift in individuals' g^*'s, i.e., the degrees of reference state inclusion at which people are just indifferent between accepting and avoiding the risk. This shift induces a tendency of increased risk acceptance, such that in the period when the new information becomes available, risk acceptance may increase to an extent exceeding the effect from "smooth" reference state inclusion. Once this "jump" is made, however, the normal risk inclusion process starts again.

Unfavorable information leads to an upward shift in individuals' g^*s, potentially resulting in a decrease in aggregated risk acceptance. If people that accepted the risk before decide to avoid it after the shift, they can induce others to avoid it due to the decrease in H. Hence, the process of adjustment to the new information can last several periods. However, once the share of risk acceptance has fully

[19] Remember from the degree of reference state inclusion in (1) that these are all individuals for whom the aggregated decisions of other members of the population matter for their reference state.

adjusted, the process of increasing risk acceptance restarts, potentially moving the share back to its former level and beyond.[20]

Similarly, an increase in the price premium π leads to a tightening of the condition for risk avoidance, i.e., an upwards shift of g^*, whereas a decrease leads to a downward shift. The resulting adjustments of the share of individuals accepting the risk are then just as described above.

Does the process ever stop for $H < 1$, i.e., before all individuals accept the risk? Consider the following intuitive explanation for the situation in period t.[21] The process continues as long as last period's increase in aggregated risk acceptance, $H_{t-1} - H_{t-2}$, induces an individual that did not accept the risk in period $t-1$ to accept it in period t. If the share of risk acceptance in the population is stable in two consecutive periods, $H_t = H_{t-1}$, it is stable for all future periods, given that there are no changes in the parameters. Hence, the self-enforcing process has ceased.

When would this happen? If one assumes the population to be continuous in the sense that g^* is continuously distributed over $[0, 1]$, the process moves smoothly on to $H = 1$. If, however, there are gaps in the distribution of g^*, the process can come to a halt. Assume that individuals are ordered according to the size of g^*_n, where $n = 1, .., N$ indexes the individual, and individual 1 is the one with the lowest g^*. Then, the process stops if there exists an individual m such that $g^*_m > g_m(H = \frac{m-1}{N})$, i.e., if, given that all individuals of lower order than individual m accept the risk, she herself still does not accept it. $H = \frac{m-1}{N}$ would then define a stationary state of the process. Changes in the parameters can make the process leave a stationary state to either a higher or a lower state H. After such changes, the process reaches a new stationary state with $H < 1$ only if there exists a gap as described above such that the state induced by the gap is of higher order (higher H) than the state that was induced by the parameter change.

Why would such gaps arise? Differences in g^*_n can be caused by, e.g., differences in individuals' utility and value functions, their reference inclusion-parameters α and ϕ, their incomes Y and their reference points r_c. A significant gap in people's g^*_ss could therefore arise if there existed clusters of consumers in the population, like high income vs. low income, quickly adapting vs. slowly adapting (e.g., technology prone vs. technology averse), highly reference-dependent preferences vs. slightly reference-dependent preferences etc. In summary, whenever the characteristics of individuals are not continuously distributed, gaps in the acceptance levels g^*_n may lead to a stable asymmetric distribution of risk acceptance ($0 < H < 1$) in spite of the risk inclusion process.

6 Implications for risk regulation

Two results of our model are particularly relevant for risk regulation. The first is derived from proposition 1 and condition (15), and was discussed in the last

[20]Note that for the individuals that accepted the risk before the new information became available, the degree of reference state inclusion is given by $\hat{g}(h = 1, H)$ in (1).

[21]A formal, evolutionary game-theoretic analysis is left for future research.

section: the process of increased risk acceptance in the population is self-enforcing. When some individuals accept the risk, this potentially leads to ever more people accepting it, without any changes in the parameters being necessary. This dynamic needs to be taken into consideration when regulating the entry of new risks into the market. Even if acceptance levels are very low at the time of entry and do not seem to merit much regulatory effort, the value that people assign to the risk may subsequently increase without that the parameters of the risk improve. But higher levels of risk acceptance induce a higher amount of risk production. If, then, the risk is judged to be harmful and to require regulation, adequate policies may be much harder to implement than prior to the market entry of the risk. Bratspies (2002) provides anecdotal evidence of this thesis. She reports that for the regulation of genetically-modified Bt corn in the US, the regulating agency (EPA) used the argument of expected slow market penetration for the initial lack of refuge requirements. With such lax regulation being in place in the beginning, subsequent implementation of tighter rules caused considerable difficulties.

The second result that is relevant for regulation is stated in Proposition 2, namely that for two identical risks, acceptance levels may differ greatly if one has a higher degree of reference state inclusion than the other. As noted in section 4.1, this may be seen as maximizing consumer welfare if the definition of welfare gives equal weight to the reference-dependent and reference-independent parts of individuals' utilities. If, instead, $v(x, r)$ is assigned less or even no importance for the determination of welfare, the risk inclusion effect leads to suboptimal decisions. A related problem from the risk regulation literature seems to support the latter view. It has been noted that for many health risks expert and lay opinions differ widely regarding the severity of these risks. For instance, consumers tend to overrate the risk induced by hazardous waste sites and underestimate the risk from indoor air pollution (see, e.g., Portney, 1992; Breyer, 1993; Hoppe and Nowak, 2004). Such differences often persist irrespectively of information campaigns. The question is then whose beliefs regulation should be based on. The trend here seems to be towards following the expert opinion, i.e., following the "rational" evaluation that is not biased by media coverage, availability heuristics, feelings of dread etc. (see Breyer, 1993; Margolis, 1996). If one agrees with this argument, one arrives at a lower weight for the "irrational" value $v(x, r)$ than for the "rational" utilities $u_p(x)$ and $u(x)$. As the result, regulatory policies would have to counterbalance the risk inclusion effect in order to avoid an "inefficient" increase in risk acceptance. In particular, regulation would have to be based on a utility measure other than people's stated preferences or WTP measures.

7 Evidence from the field

Finding evidence in support of or against the risk inclusion effect in field data is not straightforward. In most cases, habituation will work over longer periods of time, such that influences like new information about the risk or changes in media coverage are hard to distinguish from the risk inclusion effect. Hence, unambiguous evidence will be scarce. There are, however, a few instances where habituation seems to work within shorter time periods, such that other influences are relatively less important.[22] Even though such cases do not provide unambiguous evidence either, they provide some support of our thesis.

One case where risk perception changes seemed to work at accelerated speed regards the risk of developing Creutzfeld-Jakob Syndrome (CJS) after eating beef that is infected with BSE (Bovine Spongiform Encephalopathy). BSE had been known in Europe since the early nineties, but only in March 1996 was it announced by scientists in the British Parliament that BSE and CJS may be connected. The reaction of the public was immediate but short-termed in its intensity. Loy (1998) cites a study by Southey (1996) that found a reduction of beef consumption of 70% from pre-BSE levels in the week after the announcement, but of only 30-50% one month later. This means that after only 1 month of habituation, half of the reaction to the BSE-risk announcement had disappeared.

Hadjikhani et al. (1998) conducted two surveys among Swedish consumers, the first in May 1996 immediately after BSE had become known to the public through mass media coverage, and the second in August. They found that between the first and the second survey the share of consumers unwilling to buy British beef had decreased by 15-20% (from about 80% to about 60%), while the share willing to buy if it were cheaper increased by 15-25% (from almost 0 to about 20-30%). The increase in buy-if-it-is-cheaper attitudes during these three months is particularly illuminating, since it shows that for many individuals the risk is still relevant, but the WTA measure has decreased from infinity (”unwilling to buy”) to normal price levels (some share of the price for non-British beef). This is similar to the behavior observed in the insecticide experiment of Viscusi, Magat, Huber (1987). It indicates that people have got used to the risk, i.e., they partly included it in their reference states. The exact causes of this habituation process, e.g., observable behavior of others, media coverage etc. are, however, hard to infer from this study.

Pennings, Wansink, Meulenberg (2002) cite data on German consumer behavior which shows a sharp drop in beef sales after the first BSE-infected cow was found in Germany on November 26, 2000. However, already in mid-January the usual increase in beef consumption at this time of the year began, and continued to follow its usual seasonal pattern, albeit at a lower level. This suggests that at least a share of the population had got used to the risk fairly quickly.

These three examples hint to the fact that after a brief period of extreme consumption reduction, habituation set in pretty quickly and beef consumption rose again. As said above, this cannot be interpreted as unequivocal evidence of a risk inclusion effect, since other factors like intensity of media coverage and overreactions following frightening press announcements may play a role. The data are, however, supportive of our model.

Some more support is provided by a survey on the attitudes towards genetically modified food among US consumers. In general, attitudes towards gm-food in the US have been influenced by too many different factors (mainly new information and changes in the attitudes expressed in the mass media) such as to attribute trends over the last 15 years to habituation. However, the Pew Initiative on

[22]Habituation and media coverage must be expected to be closely related for some ”high publicity” risks (see the study of Hadjikhani et al., 1998). However, to keep the model clear and tractable we do not consider this link.

Food and Biotechnology published two surveys in 2001 and 2003 where subjects were first asked how safe they thought gm-foods were. After giving their answers they were confronted with the fact that more than half of the products at grocery stores already included gm-substances. Then they were asked again for their safety assessment of these products. The results were similar in both surveys: initially, 27% (29%) found them safe, while 25% found them unsafe in 2003 (2001). After being informed of their widespread use, the share of subjects believing in the products' safety increased to 44% (48%) while the sceptics dropped to 20% (21%).

We see two main reasons for the change in risk perception. First, subjects may have experienced some ad-hoc learning, inferring that if so many people eat gm-food already (including probably themselves), and no cases of adverse health effects have been reported, there will probably not be any health effects. However, such reasoning assumes that people see the health effects of gm-food as occurring in the short run. Hence, if people do not expect all effects to occur immediately, learning from others' or one's own short-term experience can explain only part of the observed effect. The second possible reason is that some kind of ad-hoc habituation took place, where the information about the widespread use of gm-food immediately changed the status quo of the subjects as to (partly) include the risk from gm-food consumption ("If so many others accept the risk already, it must be normal."). The lower risk perception would then be in line with our predictions. [23]

8 Conclusions

We have proposed a model of the inclusion of risks in people's reference states that aims at explaining the observed differences in the evaluation of common vs. new risks. The model is based on the assumption that the risk acceptance decision of an individual has an impact on other individuals that observe her decision, in particular on their getting used to this risk. Accordingly, for individuals whose preferences are only affected by their own decisions or are not reference-dependent at all, our model has no explanatory power. In this case, trends of increased risk acceptance would have to be explained by improved information, increased familiarity with the outcomes of the risk or other learning effects. Naturally, such effects will matter for individuals that comply with our assumptions, too. Hence, habituation works as one of many factors that can change the perception of risks. However, since experimental and field evidence suggest that the majority of individuals has indeed reference-dependent social-comparison preferences, habituation to risks as described in our model can be expected to have a significant influence on overall risk acceptance levels, especially if strong outside factors like vital new information are absent.

In a companion paper to this article we will consider the effect of risk inclusion if one assumes non-linear probability weights, a well established behavioral charac-

[23]Note that our model does not imply that individuals change their prior distribution over possible adverse outcomes, as may be suggested by the statement that a risk is "basically safe" or "basically unsafe". However, such statements are of relative rather than absolute nature, i.e., in comparison to other risks that the individual faces. Hence, changes in the prior and changes in the perception of the risk under a given prior can hardly be distinguished for this survey question.

teristic that we do not consider here. Besides, future research should analyze the implications of risk inclusion for monetary risks. If the source of monetary risks can be shown to be relevant for the evaluation of such risks, the risk inclusion effect may contribute to the explanation of, e.g., the equity premium puzzle. In addition, a thorough evolutionary game-theoretic analysis of the dynamic habituation process, in particular of the circumstances that lead to an interruption of the process and asymmetric stable states, could yield more detailed implications for risk regulation.

References

Abel, A. (1990), Asset Prices under Habit Formation and Catching up with the Joneses, The American Economic Review, Vol. 80, No. 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association. (May, 1990), pp. 38-42.

Barberis, N., Huang, M., and Thaler, R. (2003), Individual Preferences, Monetary Gambles and the Equity Premium, AFA 2004 San Diego Meetings.

Bratspies, Rebecca, (2002), The Illusion of Care: Regulation, Uncertainty, and Genetically Modified Food Crops, NYU Environmental Law Journal, Vol. 10, p. 297-355.

Breyer, F. and Fuchs, V. (1982), Risk Attitudes in Health: An exploratory study, NBER Working Paper No. 875.

Breyer, S. (1993), Breaking the Vicious Circle: Toward Effective Risk Regulation, Cambridge, MA: Harvard University Press.

Campbell, J. and Cochrane, J. (1999), By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, Journal of Political Economy, Vol. 107, No. 2, April 1999.

Constantinides, G. (1990), Habit Formation: A Resolution of the EEP, Journal of Political Economy 98, p. 519-543.

Coursey, D.L., Hovis, J.L. and Schulze, W.D. (1987), The Disparity between Willingness to Pay and Willingness to Accept Measures of Value, The Quarterly Journal of Economics, Vol. 102, No. 3, p. 679-690.

Duesenberry, J.S. (1949), Income, Saving, and the Theory of Consumer Behavior, Cambridge, Mass.: Harvard University Press.

Festinger, L. (1954), A theory of social comparison processes, Human Relations (7) p.117-140.

Gomes, F. and Michaelides, A. (2003), Portfolio Choice with Internal Habit Formation: A Life-Cycle Model with Uninsurable Labour Income Risk, CEPR Discussion Paper No. 3868.

Hadjikhani, A., Kaptalan-Nagy, E., Ljungren, U. and Seyed-Mohamrnad, N. (1998), Consumer Behavior and the Media: A Loosely Coupled Network, Uppsala University, Working Paper 1998/1.

Kahneman, D. and Tversky, A. (1979), Prospect Theory: An Analysis of Decision under Risk, Econometrica 47, no. 2, p. 263-292.

Kahneman, D., Knetsch, J. and Thaler, R. (1990), Experimental Tests of the Endowment Effect and the Coase Theorem, Journal of Political Economy, 98, no. 6, p. 1325-1348.

Kahneman, D., Knetsch, J. and Thaler, R. (1991), Anomalies: The Endowment Effect, Loss Aersion and Status Quo Bias, Journal of Economic Perspectives, 5, no. 1, p. 193-206.

Knetsch, J. and Sinden, J. (1984), Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value, Quarterly Journal of Economics, 99, p. 507-521.

Koszegi, B. and Rabin, M. (2004), A Model of Reference-Dependent Preferences, University of California - Berkeley, Working Paper.

List, J.A. (2003), Neoclassical Theory vs. Prospect Theory: Evidence from the Marketplace, NBER Working Paper No. 9736.

Loomes, G., Starmer, C. and Sugden, R. (2003), Do Anomalies Disappear in Repeated Markets?, The Economic Journal 113, p. C153-C166.

Loy, J.P. (1998), Auswirkungen der BSE-Krise auf die Verbraucherpreise fr Rindfleisch in Deutschland, Neununddreiigste Jahrestagung der Gesellschaft fr Wirtschafts- und Sozialwissenschaften des Landbaues vom 30. September bis 02. Oktober 1998 in Bonn.

Margolis, H. (1996), Dealing with risk: Why the public and the experts disagree on environmental issues. Chicago: University of Chicago Press.

Mellman Group and Public Opinion Strategies, The, (2001), Public Sentiment About Genetically Modified Food, for: The Pew Initiative on Food and Biotechnology.

Mellman Group and Public Opinion Strategies, The, (2003), Public Sentiment About Genetically Modified Food, for: The Pew Initiative on Food and Biotechnology.

Munro, A. and Sugden, R. (2003), On the theory of reference-dependent preferences, Journal of Economic Behavior and Organization, Vol. 50, p. 407-428.

Pennings, J.M.E., Wansink, B. and Meulenberg, M.T.G. (2002), A Note on Modeling Consumer Reactions to a Crisis: The Case of the Mad Cow Disease, International Journal of Research in Marketing, Vol. 19.

Polkovnichenko, V. (2003), Life Cycle Consumption and Portfolio Choice with Additive Habit Formation Preferences and Uninsurable Labor Income Risk, Working Paper Series.

Pollak, R.A. (1970), Habit Formation and Dynamic Demand Functions, Journal of Political Economy, Vol. 78, No. 4, p. 745-63.

Portney, P.R. (1992), Trouble in Happyville, Journal of Policy Analysis and Management, Vol. 11, No. 1, p. 131-132.

Ryder, H.E. and Heal, G.M. (1973), Optimal Growth with Intertemporally Dependent Preferences, Review of Economic Studies, Vol. 40, p. 1-33.

Samuelson, W., and Zeckhauser, R. (1988), Status Quo Bias in Decision Making, Journal of Risk and Uncertainty, 1, p. 7-59.

Savage, L. J. (1954), The Foundations of Statistics, New York: John Wiley and Sons.

Shalev, J. (2000), Loss Aversion Equilibrium, International Journal of Game Theory, 29, 2.

Slovic, P. (2001), The Perception of Risk; Risk, Society and Policy Series, Earthscan Publications Ltd., London and Sterling, VA.

Southey, C. (1996), Beef consumption well below pre BSE levels, Financial Times, London, 05.06.1996, p. 10.

Starmer, C. (2000), Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk, Journal of Economic Literature, Vol. 38, p. 332-382.

Sugden, R. (2003), Reference-dependent subjective expected utility, Journal of Economic Theory, 111, p. 172-191.

Sundaresan, S. M. (1989), Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth, Review of Financial Studies, 2, no.2: p. 73-89.

Thaler, R. (1980), Toward a Positive Theory of Consumer Choice, Journal of Economic Behavior and Organization, p. 39-60.

Thaler, R. (1999), Mental Accounting Matters, Journal of Behavioral Decision Making, 12, p. 183-206.

Viscusi, W. Kip, Magat, W. and Huber, J. (1987), An investigation of the rationality of consumer valuations of multiple health risks, RAND Journal of Economics, 18, no. 4, p. 465-479.

Wathieu, L. (1999), Consumer Habituation and Demand Patterns over Time, Working Paper.

Wood, J.V. (1989), Theory and Research Concerning Social Comparisons of Personal Attributes, Psychological Bulletin (106)2, p. 231-248.

Wu, G. and Gonzalez, R. (1999), Nonlinear Decision Weights in Choice under Uncertainty, Management Science, 45, no. 1, p. 74-85.

SFB 649 Discussion Paper Series

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Nonparametric Risk Management with Generalized Hyperbolic Distributions" by Ying Chen, Wolfgang Härdle and Seok-Oh Jeong, January 2005.
002 "Selecting Comparables for the Valuation of the European Firms" by Ingolf Dittmann and Christian Weiner, February 2005.
003 "Competitive Risk Sharing Contracts with One-sided Commitment" by Dirk Krueger and Harald Uhlig, February 2005.
004 "Value-at-Risk Calculations with Time Varying Copulae" by Enzo Giacomini and Wolfgang Härdle, February 2005.
005 "An Optimal Stopping Problem in a Diffusion-type Model with Delay" by Pavel V. Gapeev and Markus Reiß, February 2005.
006 "Conditional and Dynamic Convex Risk Measures" by Kai Detlefsen and Giacomo Scandolo, February 2005.
007 "Implied Trinomial Trees" by Pavel Čížek and Karel Komorád, February 2005.
008 "Stable Distributions" by Szymon Borak, Wolfgang Härdle and Rafal Weron, February 2005.
009 "Predicting Bankruptcy with Support Vector Machines" by Wolfgang Härdle, Rouslan A. Moro and Dorothea Schäfer, February 2005.
010 "Working with the XQC" by Wolfgang Härdle and Heiko Lehmann, February 2005.
011 "FFT Based Option Pricing" by Szymon Borak, Kai Detlefsen and Wolfgang Härdle, February 2005.
012 "Common Functional Implied Volatility Analysis" by Michal Benko and Wolfgang Härdle, February 2005.
013 "Nonparametric Productivity Analysis" by Wolfgang Härdle and Seok-Oh Jeong, March 2005.
014 "Are Eastern European Countries Catching Up? Time Series Evidence for Czech Republic, Hungary, and Poland" by Ralf Brüggemann and Carsten Trenkler, March 2005.
015 "Robust Estimation of Dimension Reduction Space" by Pavel Čížek and Wolfgang Härdle, March 2005.
016 "Common Functional Component Modelling" by Alois Kneip and Michal Benko, March 2005.
017 "A Two State Model for Noise-induced Resonance in Bistable Systems with Delay" by Markus Fischer and Peter Imkeller, March 2005.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

018 "Yxilon – a Modular Open-source Statistical Programming Language" by Sigbert Klinke, Uwe Ziegenhagen and Yuval Guri, March 2005.
019 "Arbitrage-free Smoothing of the Implied Volatility Surface" by Matthias R. Fengler, March 2005.
020 "A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics" by Matthias R. Fengler, Wolfgang Härdle and Enno Mammen, March 2005.
021 "Dynamics of State Price Densities" by Wolfgang Härdle and Zdeněk Hlávka, March 2005.
022 "DSFM fitting of Implied Volatility Surfaces" by Szymon Borak, Matthias R. Fengler and Wolfgang Härdle, March 2005.
023 "Towards a Monthly Business Cycle Chronology for the Euro Area" by Emanuel Mönch and Harald Uhlig, April 2005.
024 "Modeling the FIBOR/EURIBOR Swap Term Structure: An Empirical Approach" by Oliver Blaskowitz, Helmut Herwartz and Gonzalo de Cadenas Santiago, April 2005.
025 "Duality Theory for Optimal Investments under Model Uncertainty" by Alexander Schied and Ching-Tang Wu, April 2005.
026 "Projection Pursuit For Exploratory Supervised Classification" by Eun-Kyung Lee, Dianne Cook, Sigbert Klinke and Thomas Lumley, May 2005.
027 "Money Demand and Macroeconomic Stability Revisited" by Andreas Schabert and Christian Stoltenberg, May 2005.
028 "A Market Basket Analysis Conducted with a Multivariate Logit Model" by Yasemin Boztuğ and Lutz Hildebrandt, May 2005.
029 "Utility Duality under Additional Information: Conditional Measures versus Filtration Enlargements" by Stefan Ankirchner, May 2005.
030 "The Shannon Information of Filtrations and the Additional Logarithmic Utility of Insiders" by Stefan Ankirchner, Steffen Dereich and Peter Imkeller, May 2005.
031 "Does Temporary Agency Work Provide a Stepping Stone to Regular Employment?" by Michael Kvasnicka, May 2005.
032 "Working Time as an Investment? – The Effects of Unpaid Overtime on Wages, Promotions and Layoffs" by Silke Anger, June 2005.
033 "Notes on an Endogenous Growth Model with two Capital Stocks II: The Stochastic Case" by Dirk Bethmann, June 2005.
034 "Skill Mismatch in Equilibrium Unemployment" by Ronald Bachmann, June 2005.

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

035 "Uncovered Interest Rate Parity and the Expectations Hypothesis of the Term Structure: Empirical Results for the U.S. and Europe" by Ralf Brüggemann and Helmut Lütkepohl, April 2005.

036 "Getting Used to Risks: Reference Dependence and Risk Inclusion" by Astrid Matthey, May 2005.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".